

Projeto Semantix – Previsão de Churn em telecom

A retenção de clientes é um dos maiores desafios enfrentados pelas empresas do setor de telecomunicações. O termo “*churn*” refere-se à taxa de cancelamento de contratos por parte dos consumidores, o que gera quedas significativas na receita, aumento nos custos de aquisição e perda de vantagem competitiva. Para lidar com isso, empresas investem cada vez mais em estratégias de fidelização e relacionamento.

Clientes podem cancelar o serviço a qualquer momento, porém quais são os principais fatores que os levam a essa decisão? Atendimento insatisfatório, problemas técnicos, demora na resolução de solicitações, preços elevados ou pacotes mais atrativos da concorrência são algumas das razões mais comuns. Identificar padrões nesses comportamentos é essencial para prevenir cancelamentos e aprimorar o relacionamento com os clientes.

O churn impacta diretamente a receita da empresa, podendo gerar consequências financeiras graves e prejudicar a imagem da marca. Em casos extremos, uma taxa elevada de churn pode comprometer a viabilidade do negócio. Para empresas de telecom, cujo modelo de receita é baseado em assinaturas mensais, a fidelização dos clientes é vital para garantir crescimento sustentável.

Nesse contexto, a análise de dados surge como uma ferramenta estratégica. Técnica de machine learning permite a criação de modelos preditivos que identificam clientes com alta probabilidade de cancelamento. Isso permite que equipes de marketing atuem proativamente, oferecendo promoções, melhorias no atendimento ou planos personalizados. Com isso, a análise de dados torna-se uma aliada não apenas na retenção de clientes, como também na sustentação do negócio no longo prazo.

Fonte de dados utilizada:

Nome da base: Telco Customer Churn

Fonte: Kaggle – [Telco Customer Churn Dataset](#)

Origem: IBM (dados públicos para fins educacionais)

Descrição:

A base de dados contém informações de 7.043 clientes de uma empresa de telecomunicações dos Estados Unidos. Os dados abrangem atributos demográficos (como gênero, idade), informações de conta (como forma de pagamento e mensalidade) e dados sobre serviços contratados (como telefonia, internet, streaming, etc.). O foco da base é indicar se o cliente cancelou ou não o serviço(churn), permitindo a criação de modelos preditivos.

Tipo de dados:

Formato: CSV

Tipo de Dados: Estruturados

Catégoricos: Gender, Partner, Contract, PaymentMethod, etc.

Numéricos: tenure, MonthlyCharges, TotalCharges

Alvo: Churn

OBS: Não há dados confidenciais como CPF, nome ou endereço, o que torna a base ideal para projetos educacionais e demonstração pública.

Método de Acesso e Coleta:

Método de obtenção: Download direto via kaggle.

Formato de arquivo: CSV

Ambiente de uso:

A base foi processada em dois ambientes principais:

- Google Colab com PySpark, para leitura, limpeza de dados e análise exploratória (EDA);
- Google Planilhas, para formatação e interpretação com o Looker Studio, onde foram criadas as visualizações finais.

Análise Exploratória de Dados (EDA):

- Os dados foram analisados com PySpark no Google Colab e Google Planilhas para entender estruturas, valores ausentes e relações entre variáveis.
- Identificou-se 11 registros nulos em 'TotalCharges', todos com 'tenure sem registros', tratados com preenchimento por 0.

As variáveis numéricas foram convertidas e analisadas com '.describe()' e '.corr()', mostrando:

- Forte correlação entre 'tenure' e 'TotalCharges' (0.82);
- Correlação moderada entre 'MonthlyCharges' e 'TotalCharges' (0.65);
- Baixa correlação com 'SeniorCitizen'.

A Variável 'Churn' foi usada como alvo, com 26,5% de clientes que cancelaram.

Relatório de Insights – Projeto Churn em Telecom.

- **TotalCharges apresenta 11 valores nulos**, concentrado em clientes com tenure = 0, ou seja, recém cadastrados que ainda não geraram cobranças.
- A correlação entre as variáveis revelou:
 - TotalCharges e tenure: Possuem alta correlação (0.82).
 - MonthlyCharges e TotalCharges: Uma correlação moderada (0.65).
 - SeniorCitizen tem uma baixa correlação com qualquer variável, o que sugere baixo impacto direto, mas ainda pode ter valor preditivo combinado.
- A maioria dos clientes **não cancela** (churn = 'No'), mas há um grupo significativo que o faz (26%).

Discussão dos achados e relevância.

A alta correlação entre tenure e TotalCharges é esperada: Quanto mais tempo o cliente permanece na empresa, maior o valor acumulado em faturas. Isso indica lealdade ou uso contínuo.

A combinação de MonthlyCharges altos e tenure baixos pode sinalizar clientes insatisfeitos no início do contrato, especialmente se a percepção de custo-benefício for ruim.

A baixa correlação de SeniorCitizen pode indicar que a idade sozinha não explica o churn, mas pode ter efeito combinado com outros fatores (ex: pacote, suporte, familiaridade com tecnologia).

Algumas variáveis categóricas (como Contract, internetService, paymentMethod) mostraram, em análise descritiva, distribuições diferentes entre clientes que cancelaram e os que ficaram.

Sugestões de Ações Estratégicas.

Com base nos insights acima, são sugeridas as seguintes ações:

- Monitorar novos clientes com mensalidades altas: Esses clientes devem ser acompanhados com ofertas ou suporte personalizado, pois teriam maior chance de churn.
- Oferecer contratos de longo prazo com benefícios: Contratos do tipo “mensal” têm maior taxa de churn, incentivar planos de fidelidade (12/24 meses) pode reduzir cancelamento, podendo oferecer desconto na mensalidade (Ex: em vez de pagar uma mensalidade se pagar três meses teriam um desconto considerável que valeria a pena).
- Avaliar perfil de clientes sêniores: Embora SeniorCitizen não tenha impacto direto isolado, pode indicar necessidade de suporte especial ou plano adaptado.

Os dados confirmam que é possível prever churn com base em variáveis numéricas e categóricas bem definidas. Utilizando esses insights, a empresa pode agir de forma proativa, reduzindo perdas financeiras e aumentando a retenção de clientes com ações direcionadas.

Visualizações e Dashboard no Looker Studio.

As visualizações foram criadas no Looker Studio e organizadas em três páginas principais:

- Visão Geral – KPIs, Taxa de churn e distribuição geral.
- Perfil Cliente Churn – Análises demográficas, contratuais e de comportamento.
- Conclusão – Perfis com maior risco de cancelamento e plano de ação.

Link para acessar o dashboard:

<https://lookerstudio.google.com/reporting/a911505d-7a74-4d93-88a2-64ae7b8b43d2>